Genetic and Environmental Influences on the Relative Timing of Pubertal Change

Lindon Eaves[1], Judy Silberg[1], Debra Foley[1], Cynthia Bulik[1], Hermine Maes[1], Alaattin Erkanli[2], Adrian Angold[2], E. Jane Costello[2], and Carol Worthman[3]

[1] Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University School of Medicine, Richmond, Virginia, United States of America

[2] Epidemiology Program, Department of Psychiatry, Duke University, Durham, North Carolina, United States of America

[3] Department of Anthropology, Emory University, Atlanta, Georgia, United States of America

A multicategory item-response theory model was developed to characterize developmental changes in three items relating to the assessment of puberty in adolescent twin girls and boys. The model allowed for the fixed effects of age on probability of endorsing the responses and for the random effects of individual differences on the timing of pubertal changes relative to chronological age. In girls, the model was applied three-wave data on twin pairs ($N = 414$ female monozygotic [MZ] and 197 female dizygotic [DZ] pairs) and female twins from boy–girl pairs ($N = 300$ twins) from the Virginia Twin Study of Adolescent Behavioral Development. In boys, the data comprised 318 MZ and 185 DZ pairs and 297 male twins from boy–girl pairs. A total of 3172 and 2790 individual twin assessments were available in girls and boys, respectively, spanning ages 8–17 years. The availability of twin data allows the contributions of genes, the shared environment and individual unique environmental experiences to be resolved in the relative timing of pubertal changes. Parameters of the mixed model including fixed effects of age and random effects of genes and environment were estimated by Markov Chain Monte Carlo simulations using the BUGS algorithm for Gibbs sampling. The estimated standard deviation of random differences in the timing of puberty relative to age was 0.96 years in girls and 1.01 years in boys. The estimated intraclass correlations for the relative timing of pubertal changes were 0.99 +/–0.01 in MZ girls, 0.52 +/–0.02 in DZ girls, 0.88 +/–0.04 in MZ boys and 0.44+/–0.02 in DZ boys, indicating a very large contribution of genetic factors to the relative timing of pubertal change in both sexes. Additive genetic factors account for an estimated 96.3+/–3.3% of the total variance in random effects in girls and 88.0+/–3.6% in boys. Shared environmental influences account for 3.6+/–3.4% in girls and 0% in boys. In girls, nonshared environmental effects explain 0.1+/–0.1% of the total residual variance. The comparable figure in boys is 12.0+/–3.6%.

A large proportion of unanswered questions about the action and interaction of genetic and social factors in the development of adolescent behavior and its disorders require that we address the role of genes and environment in the underlying process of pubertal development. Thus, Silberg et al. (2001), have shown that there is a significant increase with age in the contribution of genetic factors influencing depression that may reflect the consequences of the expression of genetic differences contingent on passing the biological milestone of puberty (Eaves & Silberg, 2003). Further, Angold et al. (1999) found that the change in the rates of depression around puberty appear to correlate more highly with the latent hormonal changes accompanying puberty rather than the manifest physical changes, implying that the underlying mechanisms are biological rather than social. Meyer et al. (1997) demonstrated that measures of biological development were better predictors of changes in physical outcomes than chronological measures even though biological and chronological indices were highly correlated. Silberg et al. (2001) showed that genetic influences in depression after age 14 (when many pubertal changes have occurred) are not manifest in depression before age 12 (prior to many significant pubertal changes) but correlate almost completely with earlier genetic influences on overanxious disorder. Silberg and Bulik (2004) have argued that a similar process can be identified for the relationship between early genetic differences in anxiety and later genetic differences in eating disorder. An additional nuance to the emergence of genetic differences in depression and eating disorders after puberty is the fact that early genetic differences apparently increase exposure to later adverse life events (genotype — environment correlation) and, furthermore, increase sensitivity to the adverse impact of stressful life events (genotype x

environment interaction, Silberg et al., 2001; Silberg & Bulik, 2004; Eaves, Silberg & Erkanli, 2003)

Although the theoretical significance of the impact of genes on the timing of puberty has been recognized for the further understanding of age changes in the effects of genes and environment on psychiatric disorder (Eaves & Silberg, 2003), genetic studies of behavioral development have made very little attempt to go beyond age as a covariate of developmental change. Our future ability to analyze the action and interaction of genetic effects in behavioral development will devolve upon our ability better to characterize and separate the fixed effects of age in pubertal change from the random genetic and environmental differences in the timing of pubertal milestones that might interact with the expression of genetic effects on behavioral outcomes such as depression, eating disorders and disruptive behaviors.

We describe an approach to the analysis of the effects of genes, environment and age on pubertal changes that combines an item-response theoretical approach to the probability of passing multilevel transitions in multiple items related to puberty, with a random-effects model for variation in the timing of pubertal changes that allows for a variety of patterns of family resemblance consistent with the effects of additive genetic effects, the shared family environment and within-family environmental factors unique to members of the family (c.f. Jinks & Fulker, 1970).

The Data

We use the model to analyze pubertal changes in girls and boys aged 8–17 years from three waves of data collection from the longitudinal Virginia Twin Study of Adolescent Behavioral Development (Meyer et al., 1996; Simonoff et al., 1997; Hewitt et al., 1997; Eaves et al., 1997).

The structure of the sample of girls whose data are analyzed in this study is summarized in Table 1a. Altogether, the sample comprises observations on 1522 girls from 911 monozygotic (MZ), dizygotic (DZ) and girl–boy (DZMF) dizygotic pairs. One thousand and sixty-five individuals were measured at least twice by extended psychiatric interview using the parent and child forms of the Child and Adolescent Psychiatric Assessment (C-CAPA and P-CAPA, Angold et al., 1995; Angold & Costello, 2000). We employ five items from the pubertal assessment of the C-CAPA. The format of the semistructured pubertal assessment in both sexes is outlined in Table 2. For data analysis purposes, the response categories are coded sequentially from 1 upwards with increasing pubertal development. The number of response categories for each of the items in girls is growth spurt, 4; body hair, 4; skin changes, 3; breast development, 4; periods, 3. The comparable sample sizes for boys are provided in Table 1b. The first three items are the same in both sexes. In boys, items relating to breaking of the voice (4 response categories) and growth of facial hair (4 categories) replace those relating to breast development and onset of periods. The male sample comprises 1302 individuals from 800 pairs, of whom 964 boys were measured at least twice.

Table 3 summarizes the raw response frequencies in girls (percentage of nonmissing responses) by age in each category for each of the five items. Inspection of the raw data shows that a relatively high proportion of 16+ year-olds in the sample still do not report a

Table 1a

Structure of Sample Used for the Genetic Analysis of Pubertal Development in Girls

Source of data	Monzygotic (pairs)	Dizgyotic (pairs)	Girl–Boy (individuals)	Total (individuals)	Total (observations)
Wave 1 only	119	66	87	457	457
Waves 1 and 2	125	60	99	469	938
Waves 1,2 and 3	169	70	107	585	1755
Waves 1 and 3	1	1	6	10	20
Waves 2 and 3	0	0	1	1	2
Total	414	197	300	1522	3172

Table 1b

Structure of Sample Used for the Genetic Analysis of Pubertal Development in Boys

Source of data	Monzygotic (pairs)	Dizgyotic (pairs)	Girl–Boy (individuals)	Total (individuals)	Total (observations)
Wave 1 only	77	48	88	338	338
Waves 1 and 2	109	49	98	414	828
Waves 1,2 and 3	127	82	104	522	1566
Waves 1 and 3	5	6	7	29	58
Total	318	185	297	1302	2790

Table 2

Items Used in Assessment of Puberty (Child CAPA Self-Report)

Growth spurt

- As I expect you know, people go through a lot of physical changes as they grow up, and I want to ask you whether any of them have been happening to you?
- For instance have you been getting taller very quickly?
    - When did that start?
    - How tall are you now?
- (IF NO) Was there a time before when you were growing faster than usual?
    - When was that?

Coding: 1=No; 2=Yes, barely; 3=Yes, definitely; 4=Development complete.

Body hair

- What about body hair?
- Have you been getting any body hair – under your arms or 'down below'?
    - When did that start?
    - Is there a lot or a little?

Coding: 1=No; 2=Yes, barely; 3=Yes, definitely; 4=Development complete.

Skin changes

- Has your skin changed?
- Has it gotten more greasy?
- Has it gotten pimply?
    - When did that start?

Coding: 1=No; 2=Yes, barely; 2=Yes, definitely.

GIRLS ONLY

Breast development

- Have your breasts begun to grow?
    - When did that happen?

Coding: 1=No; 2=Yes, barely; 3=Yes, definitely; 4=Development complete.

Periods

- Have your periods started?
    - When did they start?
    - Are they regular?
    - How long have they been regular?
    - Have they ever stopped?
    - Do you have any problems with them?

Coding: 1=Not begun; 2=At least one period; 2=Regular monthly cycles established (at least 3 months).

BOYS ONLY

Voice breaking

- Has your voice broken/started to break?
    - When did that happen?

Coding: 1=No; 2=Yes, barely; 3=Yes, definitely; 4=Development complete.

Facial hair

- Have you been growing any hair on your face?
    - Has your moustache or beard started to grow?
    - When did that start?

Coding: 1=No; 2=Yes, barely; 3=Yes, definitely; 4=Development complete.

growth spurt or skin changes, suggesting that these two items are relatively weak indicators of pubertal change. By contrast, body hair, breast development and onset of periods show very few reports of no development by age 16+ and show a much steeper increase in endorsement with age. A cursory examination of the age changes suggest that, for each item, approximately 50% of the girls report some development (i.e., one of the positive response categories) as follows: growth spurt, 11–12 years of age; body hair, 10–11 years of age; skin changes, 11–12 years of age; breast development, 10–11 years of age; periods, 12–13 years of age.

Comparable figures for boys are given in Table 4. In boys, the items relating to growth spurt and skin changes appear to be more discriminating between age groups than they were in girls. The items relating to voice breaking and facial hair show relatively steep changes in endorsement frequency with age. Approximately 50% of the boys report some development (i.e., one of the positive response categories) as follows: growth spurt, 12 years of age; body hair, 11–12 years of age; skin changes, 12 years of age; voice breaking, 12–13 years of age; facial hair, 13–14 years of age. Self-reported 'growth spurt' and 'skin changes' never come close to the 100% endorsement rate expected for a pure measure of puberty and a very weak relationship with development implies that adolescents have very little grasp of the concept presented in interview format. Our subsequent model for puberty considers only the three remaining items: body hair, breast development and menarche in girls, and body hair, voice breaking and facial hair in boys.

Model

Several statistical-genetic models have already been developed for the time-dependent changes accompanying puberty. Meyer et al. (1999) proposed a survival model for menarche that incorporated random genetic and environmental differences in frailty. The model, for only one variable, was implemented using numerical optimization to maximize the likelihood of the observations. Numerical integration was performed over the range of latent genetic and environmental variables. This requirement of the ML approach renders it extremely tedious to implement because no one approach to quadrature proves equally accurate over a wide range of parameter values. Pickles et al. (1994a, 1998) developed a series of survival models for recalled ages of onset of pubertal milestones and associated behavioral events. These models all allowed for the fact that retrospective recall of the onset of past events tends to foreshorten historical time ('telescoping'), leading to biases that might be confused with effects of the shared environment. Pickles et al. (1994b) note that the apparent effects of the shared environment on pubertal development are largely removed when a well-fitting survival model is used and correction is made for

Table 3

Percentages by Age of Female VTSABD Twins in Response Categories for Five CAPA Puberty Items

Age	8	9	10	11	12	13	14	15	16	17
Growth spurt										
1	65.9	62.1	56.8	57.6	46.9	47.6	51.2	53.0	56.4	63.4
2	14.1	16.1	14.6	10.3	14.3	7.4	5.9	5.3	2.1	0.7
3	20.0	21.6	28.6	31.8	38.8	44.5	37.0	32.0	23.2	16.3
4	—	0.3	—	0.3	—	0.5	5.9	9.7	18.3	19.6
Body hair										
1	94.7	81.4	68.3	37.7	13.9	3.8	0.7	1.6	1.2	0.0
2	5.3	14.8	22.7	33.2	30.0	15.4	5.1	2.8	2.1	1.1
3	—	3.9	9.1	29.1	51.7	70.0	70.9	56.2	42.7	22.6
4	—	—	—	—	4.3	10.8	23.2	39.4	54.0	76.4
Skin changes										
1	94.0	83.9	67.7	53.8	36.6	33.8	31.1	27.9	31.9	35.3
2	6.0	13.2	26.5	33.5	36.9	31.3	30.1	25.4	24.9	22.6
3	—	2.9	5.8	12.7	26.5	34.9	38.8	46.7	43.2	42.2
Breast development										
1	85.3	73.7	51.6	31.1	13.6	2.8	2.0	1.3	1.9	0.4
2	14.7	25.0	42.9	46.0	47.5	35.4	18.1	14.4	8.2	5.2
3	—	1.3	5.5	22.9	38.3	60.0	73.8	65.6	58.7	41.0
4	—	—	—	—	0.6	1.8	6.1	18.7	31.3	53.5
Periods started										
1	100.0	99.7	97.1	88.0	57.5	32.7	8.0	2.5	0.6	0.0
2	—	0.3	1.3	4.7	8.9	9.4	6.3	2.5	3.3	1.8
3	—	—	1.6	7.3	33.6	57.9	85.6	95.0	96.1	98.2
Total	137.0	319.0	317.0	318.0	329.0	397.0	412.0	325.0	329.0	276.0

Note: Percentages include both twins and both waves and exclude missing values. Girl twins from boy–girl pairs also included in frequencies and subsequent analyses. 'Totals' are the number of individual twins assessed in each year of age (twins with observations at both waves are counted twice).

Table 4

Percentages by Age of Male VTSABD Twins in Response Categories for Five CAPA Puberty Items

Age	8	9	10	11	12	13	14	15	16	17
Growth spurt										
1	56.7	63.7	65.3	61.3	51.5	38.0	26.8	22.9	29.7	31.5
2	21.1	15.0	17.7	12.2	12.8	12.8	9.9	10.2	5.7	2.2
3	22.1	21.3	16.9	26.5	35.7	48.0	62.2	64.0	58.0	47.0
4	—	—	—	—	—	1.2	1.1	2.9	6.6	19.4
Body hair										
1	93.3	96.1	80.7	59.0	27.2	10.9	3.5	0.9	0.3	0.4
2	6.7	3.5	16.0	28.3	34.4	29.4	11.9	6.0	3.9	1.3
3	—	0.4	3.3	12.7	38.4	58.4	80.9	79.8	70.4	54.6
4	—	—	—	—	—	1.2	3.7	13.3	25.5	43.7
Skin changes										
1	98.1	95.3	82.3	69.5	52.5	38.8	29.1	31.2	30.8	27.7
2	1.9	4.7	15.3	23.4	35.7	32.3	30.7	27.1	22.2	22.9
3	—	—	2.4	7.1	11.8	28.9	40.2	41.6	47.0	49.4
Voice-breaking										
1	92.2	88.8	82.1	69.4	59.9	41.6	23.5	13.8	12.8	11.7
2	5.8	11.5	16.3	26.6	31.4	30.3	19.5	16.5	8.5	4.5
3	1.9	1.7	1.6	4.0	8.7	25.9	44.3	45.4	40.4	29.3
4	—	—	—	—	—	2.2	12.7	24.3	38.3	54.5
Facial hair										
1	98.1	96.2	91.1	85.3	68.9	56.3	33.3	15.8	5.1	2.6
2	1.9	3.8	8.9	14.7	28.0	34.4	45.1	41.5	28.7	16.6
3	—	—	—	—	3.0	9.4	21.1	40.8	61.4	69.4
4	—	—	—	—	—	—	0.5	1.9	4.8	11.4
Total	109.0	242.0	251.0	309.0	297.0	322.0	378.0	318.0	336.0	232.0

Note: Percentages include both twins and both waves and exclude missing values. Girl twins from boy–girl pairs also included in frequencies and subsequent analyses. 'Totals' are the number of individual twins assessed in each year of age (twins with observations at both waves are counted twice).

telescoping. They also note that the correlation between the random effects on frailty in MZ twins is very high, suggesting that any effects of the within-family environment are quite small.

Our model for pubertal development consists of two components: the psychometric model and the model for individual differences. The psychometric model characterizes the relationship between the items chosen to measure the construct (in this case the three 'symptoms' of pubertal development). The model for individual differences, in our case, tries to characterize the random effects of genes and environment, and the effects of covariates such as age on the latent construct of puberty. Our model thus presupposes that individuals differ for a latent trait, θ, that represents their relative position on the scale of pubertal development. The psychometric model assumes that, when θ_i is known for the ith subject, the probability of a response exceeding the kth out of n_j ordered category on the jth item is assumed to follow the cumulative logistic function:

$$P_{i,j,k} = 1/\{1+\exp[-\beta_j (\theta_i - \alpha_{j,k})]\}$$

The coefficient β_j is sometimes referred to as the 'discriminating power' of the jth item because large values of β result in a very rapid change in the probability of endorsing the item with changes in the latent trait, θ (see, e.g., Lord & Novick, 1968, for a classical statement of the logistic item-response theory [IRT] model and its assumptions). The threshold parameter, $\alpha_{j,k}$ is sometimes referred to as the 'item difficulty' because of the original connection between IRT theory and ability testing. For the jth item with n ordered response categories there are $n_j - 1$ thresholds.

Given the above cumulative probability function, the probability of endorsing the kth category of the jth item is:

$$Q_{i,j,k} = P_{i,j,k-1} - P_{i,j,k} \quad 1<k<n_j$$
$$Q_{i,j,1} = 1 - P_{i,j,1}, \quad k=1, \text{ and}$$
$$Q_{i,j,n-1} = P_{i,j,n-1}, \quad k=n_j$$

The parameters of the psychometric model are important because they allow us to evaluate the information content of the test at different parts of the scale (see Lord & Novick, 1968) and allow us, ideally, to select items that are best suited to a particular measurement goal. In our case, the thresholds will reflect the relative age at which specific transitions between response categories are most likely to occur and the discriminating powers will indicate the relative sensitivity of the items to changes in the latent trait.

In our application, the choice of items is restricted to those pubertal milestones that can be assessed noninvasively at relatively little cost in a larger face-to-face psychiatric interview. The item parameters are of interest for what they suggest about the ordering of items with respect to developmental change and the rate of change in the outcome as a function of development. However, our primary focus lies in the causes of variation in the latent dimension of puberty itself, and ultimately in our ability to delineate components of variation in pubertal development that may be used to refine subsequent analyses of developmental change in other behavioral outcomes that are known to change with age.

We hypothesize that θ_i comprises two components: a fixed effect that depends on the age of the subject and a random residual effect that reflects individual differences in the relative timing of pubertal changes. Thus:

$$\theta_i = a_i + \delta_i$$

where a_i is the age of measurement of the ith individual and δ_i the random effect on relative timing of pubertal changes in the ith individual. The units of θ are thus the units of age (years in our example) and the random effects are deviations from the values of θ predicted from age and are assumed to be N[0, σ^2]. Our model assumes that individual differences relate to the timing of pubertal changes, that is, subjects have their thresholds shifted either to the 'right' (later onset) or to the 'left' (earlier onset). We assume that the 'rates' of pubertal growth are uniform across individuals within sexes. Random effects on rates of change could, theoretically, be incorporated by allowing for variation in the sensitivity parameters, β.

In our example, the observations of puberty are not independent. Firstly, we have twin pairs so the δ_i may be correlated between MZ and DZ pairs because of the effects of genes and the shared environment on the timing of pubertal change. Secondly, for most of the pairs we have repeated measures. In our case, we assume that the repeated measures differ within individuals only by virtue of age and that the same timing difference, δ_i, applies to all ages within a subject.

Individual differences in δ may be modeled in a variety of ways. Following the classic formulation of Jinks and Fulker (1970) we model the components of variance within and between pairs of MZ and DZ twins in terms of a hierarchial random-effects model reflecting the contributions of additive genetic effects (A), the shared environment (C) and the within-family, unique environment (E). This so-called 'ACE' model is at the heart of much structural modeling of twin data (see e.g., Neale & Cardon, 1992) and has been applied extensively in our own analyses of the VTSABD data (see e.g., Eaves et al., 1997).

The total variation in random effects, σ^2 is partitioned into a component of variance between pairs, σ_b^2, and a component of variance within pairs, σ_w^2, (c.f. Jinks & Fulker, 1970; Mather & Jinks, 1982).

Further, we recognize the contribution of genetic and environmental effects is partitioned differently between and within pairs of MZ and DZ twins. Thus:

$$\sigma_{bmz}^2 = a^2 + c^2$$
$$\sigma_{wmz}^2 = e^2$$
$$\sigma_{bdz}^2 = \tfrac{1}{2}a^2 + c^2$$
$$\sigma_{wdz}^2 = \tfrac{1}{2}a^2 + e^2$$

where a^2 is the additive genetic component of variance; c^2 is the shared environmental component of variance; e^2 is the within-family, unique, environmental component. The model predicts that the total variances for MZ and DZ twins, σ_{mz}^2 and σ_{dz}^2 are equal. It is convenient to standardize the estimates of the variance components by dividing each by the estimate of the total variance: $\sigma^2 = a^2 + c^2 + e^2$, thus expressing each as a proportion of the total variance.

Other models may be fitted that include non-additive genetic effects, or rater-contrast effects but it turns out that the above model offers a good approximation to our particular data on puberty.

Although we have not conducted a formal analysis of any potential problem of model identification, the fact that the model corresponds to a constrained version of the familiar unidimensional IRT model, with a total of nine separate items (three items on three occasions) should allay most fears. The addition of twin pairs to the model does not add further complication because we have assumed that the same IRT parameters apply to both twins, and merely estimated the additional parameters to reflect the fact that twins are correlated for their values on the latent trait.

Model-Fitting

The combined IRT/age effects/random effects model was fitted to the data on MZ and DZ twins, including repeated measures where available. The model was fitted separately to data from boys and girls. Traditionally, such models have been fitted by numerical maximization of the likelihood which requires integration of the likelihood function over the (multidimensional) space of random variables (see, e.g., Eaves et al., 1987 for an example of this approach to an IRT model for twin data). This approach proves very time-consuming simply to obtain ML parameter estimates, and obtaining confidence intervals and estimates of latent trait scores (such as the random effects, δ, for individual subjects) is almost prohibitive. Obtaining the latter would be important as the relative pubertal advance or delay for individual subjects is needed for further study of the relationship between pubertal development and specific outcomes such as depression, eating disorders or externalizing behavior.

For these reasons, we chose a Markov Chain Monte Carlo approach to Bayesian estimation using the Gibbs sampler (Gilks et al., 1996). The approach generates a series of simulated samples from the distribution of the unknown parameters (components of variance, latent trait scores, IRT parameters, missing values, log-likelihoods) that converges to the population distribution of the parameters conditional upon the observed data. The approach has the advantage that after convergence ('burn-in') successive samples from the distribution of unknown parameter values may be used to construct any features of their distribution including means, standard deviations, medians and confidence intervals. These are inherent in the

MCMC approach but typically have to be programmed separately in ML approaches and require as much, if not more, computer time than is usually needed to obtain the ML estimates.

The Windows version of the program BUGS (Bayesian inference Using Gibbs Sampling, Spiegelhalter et al., 2003), WinBUGS1.3 was used to implement the MCMC approach for our example (see Do et al., 2000; Eaves & Erkanli, 2003; Eaves et al., in press, for other genetic applications). The application was written directly in the BUGS language.

Results

The estimates from the ACE model converged quite rapidly. The results are summarized for 5000 updates after an initial 6000 iteration 'burn-in'. Tables 5a and 5b summarize the psychometric parameters of the IRT model for puberty in girls and boys respectively. The threshold parameters ($\alpha_{j,k}$, in the formal model) are given as ages in years. The first threshold for each item denotes the age at which 50% of the sample make the transition from responding 'No' to each item to responding 'Yes, barely'. The thresholds correspond approximately to those that might be inferred from the raw data in Tables 3 and 4, and to the age of most rapid change in Figures 1 and 2.

Figures 1 (girls) and 2 (boys) show the item characteristic curves for the onset (first transition) for the three selected items as a function of age. The ordinate is the probability of endorsing the category 'Yes, barely' or greater. The abscissa is age. The items relating to facial and body hair and voice-breaking show the steepest item characteristic curves as a function of age.

The parameters of the ACE model for the residual effects on puberty are summarized in Tables 6a and 6b. In girls, the total variance in residual effects is $0.924+/–0.057$, corresponding to a residual standard deviation of 0.961 years. The variance in boys is $1.020+/–0.087$, yielding a standard deviation of 1.010 years. Thus, approximately 5% of adolescents have puberty advanced or delayed by as much as 1.92 years (girls) or 2.02 years (boys) on either side of the typical age of transition.

The estimated MZ correlations derived from the ACE model for relative pubertal advancement/delay are 0.999 in girls and 0.880 in boys. The correlations for DZs are 0.517 (girls) and 0.440 (boys). These statistics correspond to additive genetic factors accounting for an estimated $96.3+/–3.3\%$ of the total variance in random effects in girls and $88.0+/–3.6\%$ in boys. Shared environmental influences account for $3.6+/–3.4\%$ in girls and 0% in boys. In girls, non-shared environmental effects explain $0.1+/–0.1\%$ of the total residual variance. The comparable figure in boys is $12.0+/–3.6\%$.

Thus, in both sexes, the overwhelming source of differences in the timing of puberty relative to chronological age is genetic with significant effects of
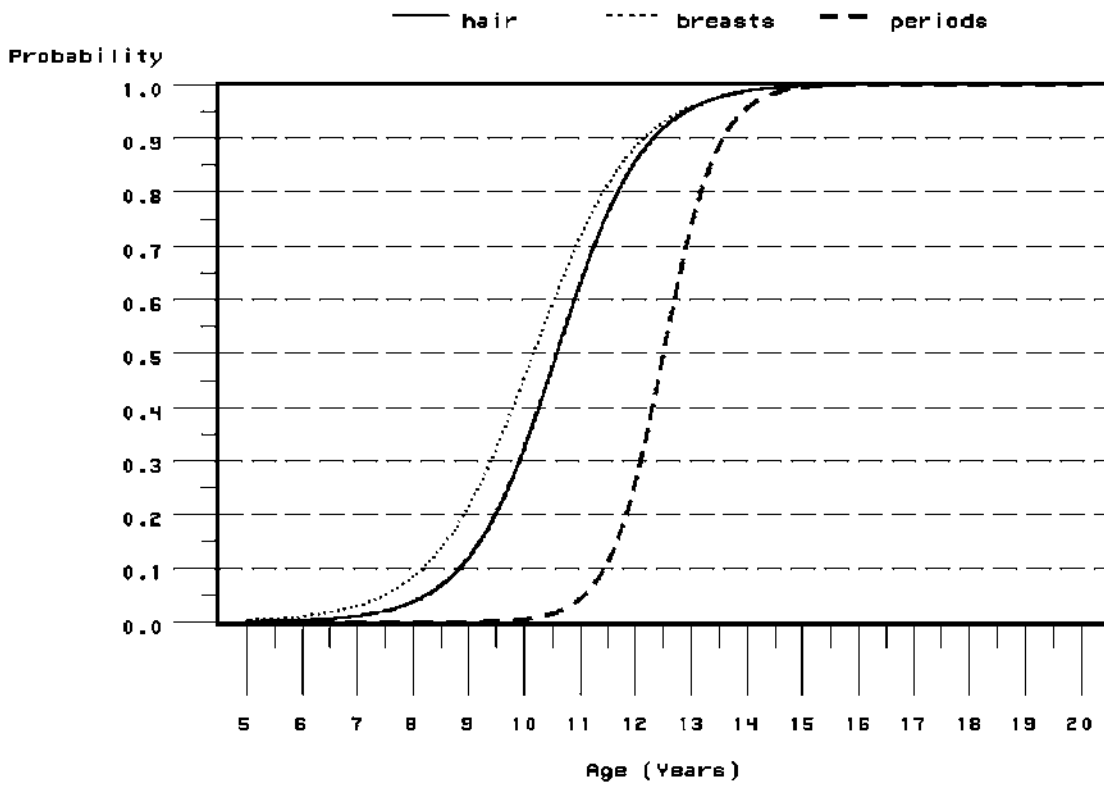
Figure 1

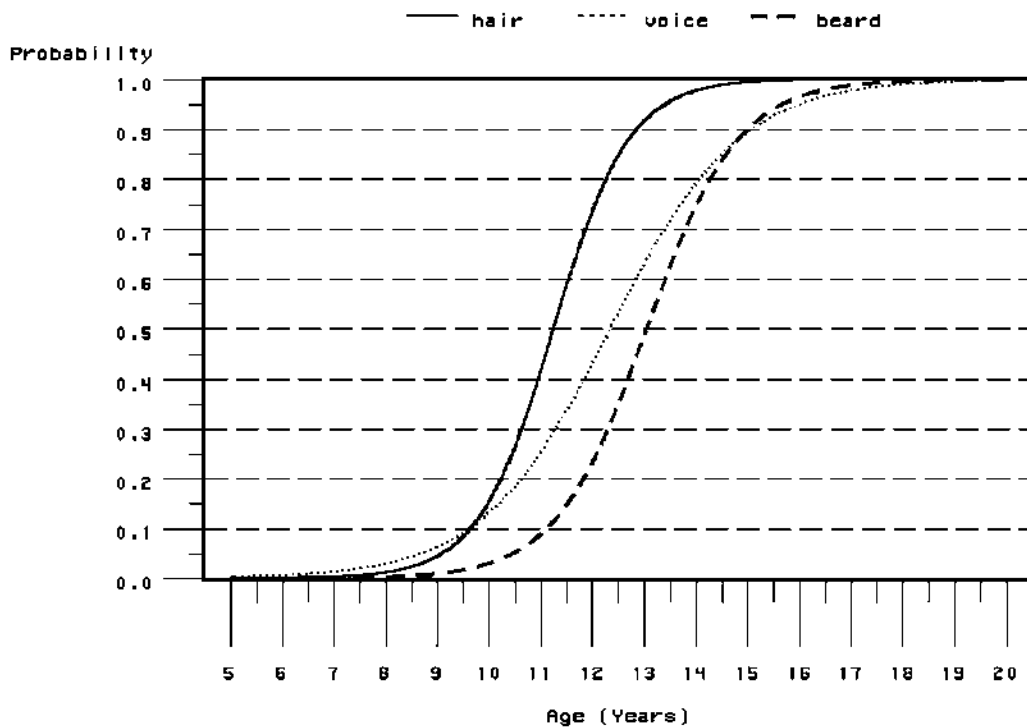Probabilities of endorsing three puberty-related items in girls from the VTSABD as a function of age.

Figure 2

Probabilities of endorsing three puberty-related items in boys from the VTSABD as a function of age.

Table 5a

MCMC Estimates of Parameters (Standard Errors in Parentheses) of Multicategory IRT Model for Pubertal Development in Girls Including Repeated Measures and MZ/DZ Twins from the VTSABD

Item	Sensitivity (s.e.)	Age threshold in years (*SE*)		
		1–2	2–3	3–4
Body hair	1.240 (0.038)	10.57 (0.058)	11.97 (0.052)	15.43 (0.067)
Breast development	1.110 (0.029)	10.17 (0.058)	12.76 (0.054)	16.70 (0.082)
Periods start	2.061 (0.091)	12.50 (0.045)	12.87 (0.044)	—

Note: Thresholds are given in years of age. Estimates based on 5000 MCMC updates after 6000 iteration 'burn in'.

Table 5b

MCMC Estimates of Parameters (Standard Errors in Parentheses) of Multicategory IRT Model for Pubertal Development in Boys Including Repeated Measures and MZ/DZ Twins from the VTSABD

Item	Sensitivity (s.e.)	Age threshold in years (*SE*)		
		1–2	2–3	3–4
Body hair	1.349 (0.046)	11.24 (0.066)	12.67 (0.060)	17.09 (0.086)
Voice breaking	0.801 (0.025)	12.35 (0.076)	14.07 (0.076)	16.83 (0.104)
Facial hair	1.139 (0.039)	13.04 (0.064)	15.37 (0.070)	19.21 (0.167)

Note: Thresholds are given in years of age. Estimates based on 5000 MCMC updates after 6000 iteration 'burn in'.

Table 6a

Parameters of Random Effects Model for Latent Distribution of Residual Pubertal Effects (i.e., Advancement/Delay Relative to Age) in MZ and DZ Girls

Statistic	Mean	*SD*	2.5 %-ile	Median	97.5 %-ile
Total variance	0.924	0.057	0.818	0.923	1.036
% Additive genetic	96.3	3.3	89.1	97.2	99.7
% Shared environment	3.6	3.4	0.	2.8	10.9
% Nonshared environment	0.1	0.1	0.	0.1	0.5
MZ correlation	0.999	0.001	0.995	0.999	1.000
DZ correlation	0.517	0.017	0.498	0.514	0.554
2 ln(l)	11190.0	55.07	11080.0	11190.0	11290.0

Note: Estimates based on sample of 5000 MCMC iterations after 6000 iteration 'burn in'.

Table 6b

Parameters of Random Effects Model for Latent Distribution of Residual Pubertal Effects (i.e., Advancement/Delay Relative to Age) in MZ and DZ Boys

Statistic	Mean	s.d.	2.5%-ile	Median	97.5%-ile
Total Variance	1.020	0.087	0.863	1.017	1.201
% Additive Genetic	88.0	3.6	80.3	88.2	95.8
% Shared Environment	0	—	—	—	—
% Nonshared environment	12.0	3.6	4.2	11.8	19.7
MZ correlation	0.880	0.036	0.803	0.882	0.958
DZ correlation	0.440	0.018	0.402	0.441	0.479
−2 ln(l)	12030.0	51.55	11930.0	12030.0	12130.0

Note: Estimates based on sample of 5000 MCMC iterations after 6000 iteration 'burn in'.

Table 7

Illustrative MCMC Estimates of Individual Random Effects in MZ Twin Pairs on the Advancement or Delay in Puberty Relative to Age

Illustrative MZ twin scores for girls				Illustrative MZ twin scores for boys			
Twin 1 score	Error	Twin 2 score	Error	Twin 1 score	Error	Twin 2 score	Error
−2.560	0.165	−2.559	0.166	−2.548	0.483	−2.364	0.487
−2.051	0.231	−2.049	0.230	—			
−1.542	0.179	−1.543	0.178	−1.646	0.506	−1.548	0.504
−0.970	0.276	−0.969	0.178	−1.005	0.570	−0.951	0.583
−0.537	0.413	−0.539	0.412	−0.471	0.963	−0.514	0.974
−0.013	0.193	−0.014	0.194	−0.035	0.498	−0.001	0.495
0.484	0.279	0.484	0.277	0.514	0.487	0.646	0.490
0.998	0.257	0.998	0.257	1.092	0.731	0.944	0.707
1.499	0.266	1.500	0.265	1.502	0.630	1.491	0.613
2.021	0.320	2.019	0.320	2.056	0.434	1.592	0.441
2.449	0.195	2.449	0.195	2.352	0.488	2.355	0.509

Note: Estimates are based on 1000 iterations after an 11,000 iteration 'burn-in' of the IRT 'ACE' model.

the shared environment in girls and significant effects of the nonshared environment that contribute, for example, to differences within MZ pairs in the relative timing of puberty. Based on these figures, 95% of all female MZ twin pairs are expected to have onsets of puberty within only 0.86 months of each other compared with as much as nearly two years difference between individuals paired randomly from the population. In boys, the comparable differences are 8.35 months for MZ twins.

The fact that the estimates of E are small in both sexes, almost zero in girls, does not mean that environmental differences within families have no effect. The E in our model only reflects environmental influences that influence the latent trait across all items. Item-specific environmental effects are reflected in the slopes, β, of the item-response curves. Small values of β correspond to large stochastic effects on item response contingent on the latent trait, or item-specific environmental effects.

An added bonus of the MCMC approach is the fact that it can provide at relatively little cost estimates of the individual random effects, δ. These estimates may be used as covariates in analyses of the effects of puberty on other outcome variables. Table 7 gives illustrative estimates of the random effects with standard errors for samples of MZ female and male twin pairs across the range of values of δ. The estimates are based on 1000 updates after the 11,000 iteration 'burn-in'. The errors of the estimated scores are large, but there is no marked tendency for errors to be greater in the extremes than in the middle or vice versa. The tabulated data, although they were chosen to represent the range of scores, were not specifically chosen for particular similarity or dissimilarity between MZ twins. They thus illustrate the remarkable concordance in MZ twins for the ranking of pubertal development relative to age.

Discussion

Our analysis shows that it is possible to combine an IRT model for the fixed and random effects of pubertal change in multiple possible indices of puberty with a model for the genetic and environmental effects required to account for patterns in family resemblance for the random departures from the fixed effects of chronological age on puberty. The MCMC approach provides a convenient platform for the statistical analysis that yields a number of additional features that are only available at much greater computational cost, if at all, through the conventional approach of maximum likelihood. These advantages include: estimates of the standard errors and confidence intervals of model parameters and functions of parameters, and estimates of individual random effects with their standard errors. In this case, the approach allows us to assign to each individual subject scores that reflect their advancement or delay of pubertal development relative to peers in their age group. We find that random departures from what is typical for age are due largely to additive genetic factors, with modest but significant effects of the shared environment (in girls) and influences of the nonshared environment (in both sexes). None of these effects can be discounted in the model. The analysis only employed age to the nearest year. Residual age changes within a year may explain some of the apparent effects of the shared environment.

Our picture of the relative timing of major pubertal changes obtained by self-report in a semi-structured psychiatric interview compares quite favorably with those obtained by more direct, but intrusive, assessment. Three other smaller twin studies (see review by Treloar & Martin, 1990) give average correlations for age at menarche of 0.93 for MZ and 0.62 for DZ twins. Our analysis does not yield direct estimates of the twin correlations for individual items related to

puberty, but polychoric correlations for the categorical menarche item in our sample are 0.97, 0.91, 0.86 and 0.82 for MZ pairs aged 11, 12, 13 and 14 years of age at interview respectively (N = 77, 80, 114 and 112 pairs). The corresponding polychoric correlations for DZ pairs are 0.51, 0.54, 0.67 and 0.84 (N = 41, 51, 47 and 54 pairs). The unweighted average of these correlations is 0.89 for MZ and 0.64 for DZ pairs – remarkably close to the correlations derived from older published studies. A remarkable exception to this pattern is that reported for retrospective questionnaire recall of age of menarche in a much larger sample of adult Australian twins (Treloar & Martin, 1990). In this sample the overall MZ correlation was 0.65 (N = 1177 pairs) and the DZ correlation was 0.18 (N = 711). The standard deviation of age at menarche is approximately 1.2, slightly larger than our estimated standard deviation for pubertal age in girls. Not only are these correlations much smaller than those reported by other authors, the very low DZ correlation relative to MZs in the Australian sample has been offered as evidence of nonadditive genetic variation in menarche, characteristic of a trait showing a strong relationship to fitness. The reason for the discrepancy is not clear. The test–test reliability of the Australian data is 0.91. Correction for this source of unreliability would increase the MZ and DZ correlations to 0.78 and 0.22 respectively and fails to resolve the apparent discrepancy. In the absence of empirical data on the effects of recall, further attempts to resolve the difference are purely speculative.

We anticipate that the basic approach we have outlined in this paper may, in the future, be integrated with analyses of the role of pubertal influences on the expression of genetic and social factors on differences in behavioral outcomes that are already known to occur but for which clearly articulated analytical strategies, grounded in a statistical genetic theory, have yet to be devised.

Acknowledgments

References

Angold, A., & Costello, E. J. (2000). The Child and Adolescent Psychiatric Assessment (CAPA). *Journal of the American Academy of Child and Adolescent Psychiatry, 39,* 39–48.

Angold, A., Costello, E. J., Erkanli, A., & Worthman, C. M. (1999). Pubertal changes in hormone levels and depression in girls. *Psychological Medicine, 29,* 1043–1053.

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E., & Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA). *Psychological Medicine, 25,*739–753.

Do, K. A., Broom, B. M., Kuhnert, P., Duffy, D. L., Todorov, A. A., Treloar, S. A., & Martin, N. G. (2000). Genetic analysis of the age at menopause by using estimating equations and Bayesian random effects models. *Statisitics in Medicine, 19,*1217–1235.

Eaves, L. J., Erkanli, A., Silberg, J., Angold, A., Maes, H. H., & Foley, D. (in press). Application of Bayesian Inference using Gibbs Sampling ('BUGS') to item-response theory modeling of multi-symptom genetic data. *Behavior Genetics.*

Eaves, L. J., & Erkanli, A. (2003). Markov Chain Monte Carlo approaches to analysis of genetic and environmental components of human developmental change and GxE interaction. *Behavior Genetics, 33,* 279–299.

Eaves, L. J., & Silberg, J. L. (2003). Modulation of gene expression by genetic and environmental heterogeneity in the timing of a developmental milestone. *Behavior Genetics, 33,* 1–6.

Eaves, L. J., Martin, N. G., Heath, A. C., & Kendler, K. S. (1987). Testing genetic models for multiple symptoms: an application to the genetic analysis of liability to depression. *Behavior Genetics, 17,* 331–341.

Eaves, L. J., Silberg, J. L., & Erkanli, A. (2003). Resolving multiple epigenetic pathways to adolescent depression. *Journal of Child Psychology and Psychiatry, 44,* 1006–1014.

Eaves, L. J., Silberg, J. L., Meyer, J. M., Maes, H. H., Simonoff, E., Pickles, A., Rutter, M., Neale, M. C., Reynolds, C., Erickson, M., Heath, A., Loeber, R., Truett, K. R., & Hewitt, J. K. (1997). Genetics and developmental psychopathology: 2. The main effects of genes and environment on behavioral problems in the Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child Psychology and Psychiatry, 38,* 965–980.

Gilks, W. R., Richardson, S., & Spielgelhalter, D. (1996). *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

Hewitt, J. K., Silberg, J. L., Rutter, M. L., Simonoff, E., Meyer, J. M., Maes, H. H., Pickles, A. R., Neale, M. C., Loeber, R., Erickson, M. T., Kendler, K. S., Heath, A. C., Truett, K. R., Reynolds, C. A., & Eaves, L. J. (1997). Genetics and developmental psychopathology: 1. Phenotypic assessment in The Virginia Twin Study of Adolescent Behavioral Development. *Journal of Child Psychology and Psychiatry, 38,* 943–963.

Jinks, J. L., & Fulker, D. W. (1970). Comparison of the biometrical genetical, MAVA and classical approaches to the analysis of human behavior. *Psychological Bulletin, 73,* 311–349.

Lord, F. M., & Novick, M. R. (1968). *Statistical Theories of Mental Test Scores.* New York: Addison-Wesley.

Mather, K., & Jinks, J. L. (1982). *Biometrical genetics: The study of continuous variation.* London: Chapman and Hall.

Meyer, J. M., Silberg, J. L., Eaves, L. J., Maes, H. H., Simonoff, E., Pickles, A., Rutter, M. L., & Hewitt, J. K. (1999). Variable age of gene expression: implications for developmental genetic models. In M. C. LaBuda, & E. L. Grigorenko (Eds.), *On the Way to Individuality: Current Methodological Issues in Behavioral Genetics* (pp. 23–52). Moscow, Russia: Progress.

Meyer, J. M., Silberg, J. L., Simonoff, E., Kendler, K. S., & Hewitt, J. K. (1996). The Virginia Twin Study of Adolescent Behavioral Development: Assessing sample biases in demographic correlates of psychopathology. *Psychological Medicine, 26,* 1119–1133.

Neale, M. C., & Cardon, L. (1992). *Methodology for Genetic Studies of Twins and Families.* Dodrecht, The Netherlands: Kluwer Academic.

Pickles, A. R., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., Hewitt, J., & Silberg, J. (1994a). Survival models for developmental genetic data: Age of onset of puberty and antisocial behavior in twins. *Genetic Epidemiology, 11,* 155–170.

Pickles, A. R., Crouchley, R., Simonoff, E., Eaves, L., Meyer, J., Rutter, M., Hewitt, J., & Silberg, J. (1998). Genetic 'clocks' and 'soft' events: A twin model for pubertal development and other recalled sequences of developmental milestones, transitions, or ages at onset. *Behavior Genetics, 28,* 243–53.

Pickles, A., Neale, M. C., Simonoff, E., Rutter, M., Hewitt, J. K., Meyer, J. M., Crouchly, R., Silberg, J. L., & Eaves, L. J. (1994b). A simple method for censored age–of–onset data subject to recall bias: Mothers' reports of age of puberty in male twins. *Behavior Genetics, 24,* 457–468.

Silberg, J., & Bulik, C. M. (2004). *The developmental association between eating disorders symptoms and symptoms of depression and anxiety in juvenile twin girls.* Manuscript submitted for publication.

Silberg, J., Rutter, M., & Eaves, L. J. (2001). Genetic and environmental influences on the temporal association between early anxiety and later depression in girls. *Biological Psychiatry, 49,* 1040–1049.

Simonoff, E., Pickles, A. R., Meyer, J. M., Silberg, J. L., Maes, H. H., Loeber, R., Rutter M. L., Hewitt, J. K., & Eaves, L. J. (1997). The Virginia Twin Study of Adolescent Behavioral Development: Influences of age, gender and impairment on rates of disorder. *Archives of General Psychiatry, 54,* 801–808.

Spiegelhalter, D. J., Thomas, A., & Best, N.G. (2003). *WinBUGS version 1.4 User Manual.* Cambridge, UK: MRC Biostatistics Unit.

Treloar, S., & Martin, N. G. (1990). Age at menarche as a fitness trait: non-additive genetic variance detected in a large twin sample. *American Journal of Human Genetics, 47,* 137–148.